



Identifying Effects of Multiple Treatments in the Presence of Unmeasured Confounding

Wang Miao, Wenjie Hu, Elizabeth L. Ogburn & Xiao-Hua Zhou

To cite this article: Wang Miao, Wenjie Hu, Elizabeth L. Ogburn & Xiao-Hua Zhou (2022): Identifying Effects of Multiple Treatments in the Presence of Unmeasured Confounding, Journal of the American Statistical Association, DOI: [10.1080/01621459.2021.2023551](https://doi.org/10.1080/01621459.2021.2023551)

To link to this article: <https://doi.org/10.1080/01621459.2021.2023551>

 View supplementary material [↗](#)

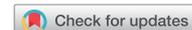
 Published online: 28 Feb 2022.

 Submit your article to this journal [↗](#)

 Article views: 1440

 View related articles [↗](#)

 View Crossmark data [↗](#)



Identifying Effects of Multiple Treatments in the Presence of Unmeasured Confounding

Wang Miao^a , Wenjie Hu^a, Elizabeth L. Ogburn^b , and Xiao-Hua Zhou^c 

^aDepartment of Probability and Statistics, Peking University, Beijing, PR China; ^bDepartment of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; ^cDepartment of Biostatistics and Beijing International Center for Mathematical Research, Peking University, Beijing, PR China

ABSTRACT

Identification of treatment effects in the presence of unmeasured confounding is a persistent problem in the social, biological, and medical sciences. The problem of unmeasured confounding in settings with multiple treatments is most common in statistical genetics and bioinformatics settings, where researchers have developed many successful statistical strategies without engaging deeply with the causal aspects of the problem. Recently there have been a number of attempts to bridge the gap between these statistical approaches and causal inference, but these attempts have either been shown to be flawed or have relied on fully parametric assumptions. In this article, we propose two strategies for identifying and estimating causal effects of multiple treatments in the presence of unmeasured confounding. The *auxiliary variables* approach leverages variables that are not causally associated with the outcome; in the case of a univariate confounder, our method only requires one auxiliary variable, unlike existing instrumental variable methods that would require as many instruments as there are treatments. An alternative *null treatments* approach relies on the assumption that at least half of the confounded treatments have no causal effect on the outcome, but does not require a priori knowledge of which treatments are null. Our identification strategies do not impose parametric assumptions on the outcome model and do not rest on estimation of the confounder. This article extends and generalizes existing work on unmeasured confounding with a single treatment and models commonly used in bioinformatics. Supplementary materials for this article are available online.

ARTICLE HISTORY

Received November 2020
Accepted December 2021

KEYWORDS

Confounding; Identification;
Instrumental variable;
Multiple treatments

1. Introduction

Identification of treatment effects in the presence of unmeasured confounding is a persistent problem in the social, biological, and medical sciences, where in many settings it is difficult to collect data on all possible treatment-outcome confounders. Identification means that the treatment effect of interest is uniquely determined from the joint distribution of observed variables. Without identification, statistical inference may be misleading and is of limited interest. Most of the work on unmeasured confounding by causal inference researchers focuses on settings with a single treatment, and either harnesses auxiliary variables (e.g., instruments, negative controls, or confounder proxies) to achieve point identification of causal effects, or relies on sensitivity analyses or on weak assumptions to derive bounds for the effects of interest. A large body of work from statistical genetics and computational biology is concerned with multiple treatments – for example Genome-Wide Association Studies (GWAS) with confounding by population structure and computational biology applications with confounding by batch effects. Recently, there have been a few attempts to put these approaches on solid theoretical footing and to bridge the gap between these statistical approaches and causal inference. However, these attempts either rely themselves on strong parametric models that circumvent the underlying causal structure, or have

been shown to be flawed. In this article, we propose two novel strategies for identifying causal effects of multiple treatments in the presence of unmeasured confounding without placing any parametric restrictions on the outcome model. This article generalizes existing work on unmeasured confounding with a single treatment to the multi-treatment setting, and resolves challenges that have undermined previous proposals for dealing with multi-treatment unmeasured confounding.

1.1. Related Work

For a single treatment, a variety of methods have been developed to test, adjust for, and eliminate unmeasured confounding bias. Sensitivity analysis (Cornfield et al. 1959; Rosenbaum and Rubin 1983; Ding and Vanderweele 2014) and bounding (Manski 1990; Balke and Pearl 1997; Richardson and Robins 2014) are used to evaluate the robustness of causal inference to unmeasured confounding. For point identification of the treatment effect, the instrumental variable (IV) is an influential tool used in biomedical, epidemiological, and socioeconomic studies (Wright 1928; Goldberger 1972; Robins 1994; Angrist, Imbens, and Rubin 1996; Didelez and Sheehan 2007; Small et al. 2017). Recently, Miao, Geng, and Tchetgen Tchetgen (2018), Shi et al. (2020), Tchetgen Tchetgen et al. (2020), Lipsitch, Tchetgen Tchetgen, and Cohen (2010), Kuroki and Pearl,

CONTACT Wang Miao  mwfy@pku.edu.cn  Department of Probability and Statistics, Peking University, Beijing, PRC. Xiao-Hua Zhou  azhou@math.pku.edu.cn

 Beijing International Center for Mathematical Research and Department of Biostatistics, Peking University, Beijing, PRC

 Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

(2014) Ogburn and VanderWeele (2012), Flanders, Strickland, and Klein (2017), and Wang et al. (2017) demonstrate the potential of using confounder proxy variables and negative controls for adjustment of confounding bias. For an overview of recent work in the single treatment setting, see Tchetgen Tchetgen et al. (2020) and Wang and Tchetgen Tchetgen (2018).

Similar methods can sometimes be used in settings with multiple treatments, simply treating them as a single vector-valued treatment. These approaches allow for unrestricted correlations among the treatments. However, if, as is typically the case in GWAS and computational biology settings, correlations among treatments contains useful information about the confounding, these methods cannot leverage the information. Latent variable methods leveraging the multi-treatment correlation structure have been used to estimate and control for unmeasured confounders in biological applications since the early 2000s (Alter, Brown, and Botstein 2000; Price et al. 2006; Leek and Storey 2007; Friguet, Kloareg, and Causeur 2009; Gagnon-Bartsch and Speed 2012; Luo and Wei 2019). Recently, a few authors have attempted to elucidate the causal structure underlying these statistical procedures and to establish rigorous theoretical guarantees for identification, using fully parametric models. Wang et al. (2017) propose confounding adjustment approaches for the effects of a treatment on multiple outcomes under a linear factor model; by reversing the labeling of the outcome and treatments, their approaches can test but not identify the effects of multiple treatments on the outcome. Kong, Yang, and Wang (2022) consider a binary outcome with a univariate confounder and prove identification under a linear factor model for multiple treatments and a parametric outcome model via a meticulous analysis of the link distribution; but their approach cannot generalize to the multivariate confounder setting as we illustrate with a counterexample in the supplement. Grimmer, Knox, and Stewart (2020), Čevd, Bühlmann, and Meinshausen (2020), Guo, Čevd, and Bühlmann (2020), and Chernozhukov et al. (2017) consider linear outcome models with high-dimensional treatments that are confounded or mismeasured; in this case, identification is implied by the fact that confounding on each treatment vanishes as the number of treatments goes to infinity. In contrast, we take a fundamentally causal approach to confounding and to identification of treatment effects by allowing the outcome model to be unrestricted, the treatment-confounder distribution to lie in a more general, though not unrestricted, class of models, the number of treatments to be finite, and confounding to not vanish.

Most notably, Wang and Blei (2019) provide an intuitive justification for using latent variable methods in general multi-treatment unmeasured confounding settings; they call the justification and resulting method the “deconfounder.” Their approach uses a factor model assuming that treatments are independent conditional on the confounder to estimate the confounder, and the confounder estimate is used for adjustment of bias. However, as demonstrated in a counterexample by D’Amour (2019a) and discussed by Ogburn, Shpitser, and Tchetgen Tchetgen (2019, 2020) and Imai and Jiang (2019), identification is not guaranteed for the deconfounder, that is, the treatment effects cannot be uniquely determined from the observed data even with an infinite number of data samples. Additionally, an infinite number of treatments are required for

consistent estimation of the confounder, complicating finite sample inference and undermining positivity. For refinements and discussions of the deconfounder approach, see Wang and Blei (2020), D’Amour (2019a), Ogburn, Shpitser, and Tchetgen Tchetgen (2020), Grimmer, Knox, and Stewart (2020), and the commentaries (D’Amour 2019b; Ogburn, Shpitser, and Tchetgen Tchetgen 2019; Imai and Jiang 2019; Athey, Imbens, and Pollmann 2019) published alongside Wang and Blei (2019). D’Amour (2019a) suggests the proximal inference and Imai and Jiang (2019) consider the conventional instrumental variable approach to facilitate identification. However, if correlations among the multiple treatments are indicative of confounding, as the deconfounder approach assumes, neither of these two methods makes use of that correlation. Moreover, their extension to the multi-treatment setting is complicated by the fact that the proximal inference requires confounder proxies to be causally uncorrelated with any of the treatments and the instrumental variable approach requires at least as many instrumental variables as there are treatments.

1.2. Contribution

In Section 2, we review the challenges for identifying multi-treatment effects in the presence of unmeasured confounding. In Sections 3 and 4, we propose two novel approaches for the identification of causal effects of multiple treatments with unmeasured confounding: an *auxiliary variables* approach and a *null treatments* approach. Both approaches rely on two assumptions restricting the joint distribution of the unmeasured confounder and treatments. The first assumption is that the joint treatment-confounder distribution lies in a class of models that satisfy a particular *equivalence* property that is known to hold for many commonly used models, for example, many types of factor and mixture models. This assumption can accommodate other treatment-confounder models, such as mixture models, in addition to the factor models considered by Wang and Blei (2019), Kong, Yang, and Wang (2022), Wang et al. (2017), and Grimmer, Knox, and Stewart (2020). The second assumption is that the treatment-confounder distribution satisfies a *completeness* condition that is standard in nonparametric identification problems. In addition to these two assumptions, the auxiliary variables approach leverages an auxiliary variable that does not directly affect the outcome to identify treatment effects, such as an IV or confounder proxy. In the presence of a univariate confounder, identification can be achieved with our approach even if only one auxiliary variable is available and if it is associated with only one confounded treatment. In contrast, IV approaches require as many instrumental variables as there are treatments and that all confounded treatments must be associated with the instrumental variables. The null treatments approach does not require any auxiliary variables, but instead rests on the assumption that at least half of the confounded treatments are null, without requiring knowledge of which are active and which are null. In these two approaches, identification is achieved without imposing parametric assumptions on the outcome model, although the joint treatment-confounder distribution is restricted by the equivalence and the completeness assumptions. Identification does not rest on estimation

of the unmeasured confounder, and thus, works with a finite number of treatments and does not run afoul of positivity. In the absence of auxiliary variables and if the null treatments assumption fails to hold, our method still constitutes a valid test of the null hypothesis of no joint treatment effect. Because identification in both approaches requires solving an integral equation, an explicit identification formula is not available for unrestricted outcome models. However, we describe some estimation strategies in Section 5. In simulations in Section 6, the proposed approaches perform well with little bias and appropriate coverage rates. In a data example about mouse obesity in Section 7, we apply the approaches to detect genes possibly causing mouse obesity, which reinforces previous findings by taking unmeasured confounding into account. Section 8 concludes with a brief mention of some potential extensions of our approaches. Proofs and further discussions are relegated to the supplementary material.

2. Preliminaries and Challenges to Identification

Throughout the article, we let $X = (X_1, \dots, X_p)^T$ denote a vector of p treatments and Y an outcome. We are interested in the effects of X on Y , which may be confounded by a vector of q unobserved covariates U . The dimension of the confounder, q , is assumed to be known a priori; for choice of q in practice see illustrations in Sections 6–7 and the discussion in Section 8. For notational convenience, we suppress observed covariates, and all conditions and results can be viewed as conditioning on them. Hereafter, we use Σ_A to denote the covariance matrix of a random vector A . We use f to denote a generic probability density or mass function and $f(A = a | B = b)$ the conditional density/mass of A given B evaluated at $(A = a, B = b)$, and write $f(a | b)$ for simplicity. Vectors are assumed to be column vectors unless explicitly transposed. We refer to $f(x, u)$ as the treatment-confounder distribution and $f(y | u, x)$ the outcome model.

Let $Y(x)$ denote the potential outcome that would have been observed had the treatment X been set to x . Treatment effects are defined by contrasts of potential outcomes between different treatment conditions, and thus, we focus on identification of $f\{Y(x)\}$. We say that $f\{Y(x)\}$ is identified if and only if it is uniquely determined by the joint distribution of observed variables.

Throughout we make three standard identifying assumptions.

- Assumption 1.* (i) Consistency: When $X = x, Y = Y(x)$;
- (ii) Ignorability: $Y(x) \perp\!\!\!\perp X | U$;
- (iii) Positivity: $0 < f(X = x | U = u) < 1$ for all (x, u) .

Consistency states that the observed outcome is a realization of the potential outcome under the treatment actually received. Ignorability, also called “exchangeability,” ensures that treatment assignments are effectively randomized conditional on U and implies that U suffices to control for all confounding. Positivity, also called “overlap,” ensures that for all values of U all treatment values have positive probability.

If we were able to observe the confounder U , Assumption 1 would permit fully nonparametric identification of $f\{Y(x)\}$ by

the back-door formula (Pearl 1995) or the g-formula (Robins 1986),

$$f\{Y(x) = y\} = \int_u f(y | u, x)f(u)du. \tag{1}$$

But when U is not observed, all information contained in the observed data is captured by $f(y, x)$, from which one cannot uniquely determine the joint distribution $f(y, x, u)$. To be specific, one has to solve for $f(x, u)$ and $f(y | u, x)$ from

$$f(x) = \int_u f(x, u)du, \tag{2}$$

$$f(y | x) = \int_u f(y | u, x)f(u | x)du. \tag{3}$$

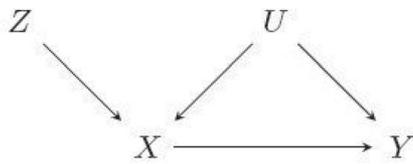
However, $f(x, u)$ cannot be uniquely determined from (2), even if, as is common practice, a factor model is imposed on $f(x, u)$; see D’Amour (2019a) for a counterexample. Furthermore, even if $f(x, u)$ is known, the outcome model $f(y | u, x)$ cannot be identified; D’Amour (2019a) points out that lack of identification of $f(y | u, x)$ is due to the unknown copula of $f(y | x)$ and $f(u | x)$. Here we note that identifying $f(y | u, x)$ given $f(u | x)$ is equivalent to solving the integral equation (3), but the solution is not unique without extra assumptions. As a result, one cannot identify the true joint distribution $f(y, x, u)$ that is essential for the g-formula (1). We call a joint distribution $\tilde{f}(y, x, u)$ *admissible* if it conforms to the observed data distribution $f(y, x)$, that is, $f(y, x) = \int_u \tilde{f}(y, x, u)du$. The counterexample by D’Amour (2019a) also shows that different admissible joint distributions result in different potential outcome distributions, that is, the potential outcome distribution is not identified without additional assumptions. Some previous approaches have estimated U directly with a deterministic function of X , but this controverts the positivity assumption and requires an infinite number of treatments in order to consistently estimate U . Furthermore, Grimmer, Knox, and Stewart (2020) show that in these settings the effect of X on Y is asymptotically unconfounded and a naive regression of Y on X weakly dominates these more involved approaches.

3. Identification with Auxiliary Variables

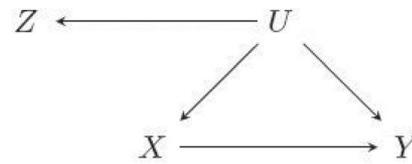
3.1. The Auxiliary Variables Assumption

Suppose we have available a vector of auxiliary variables Z , then the observed data distribution is captured by $f(x, y, z)$, from which we aim to identify the potential outcome distribution $f\{Y(x)\}$. Let $f(x, u | z; \alpha)$ denote a model for the treatment-confounder distribution indexed by a possibly infinite-dimensional parameter α , and $f(x | z; \alpha)$ the resulting marginal distribution. Given $f(x | z; \alpha)$, we let $f(x, u | z; \tilde{\alpha})$ denote an arbitrary admissible joint distribution such that $f(x | z; \alpha) = \int_u f(x, u | z; \tilde{\alpha})du$, and write $\tilde{f}(x, u | z) = f(x, u | z; \tilde{\alpha})$ for short. Our identification strategy rests on the following assumption.

- Assumption 2.* (i) Exclusion restriction: $Z \perp\!\!\!\perp Y | (X, U)$;
- (ii) Equivalence: for any α , any $\tilde{f}(x, u | z)$ that solves $f(x | z; \alpha) = \int_u \tilde{f}(x, u | z)du$ can be written as $\tilde{f}(x, u | z) = f\{X = x, V(U) = u | z; \alpha\}$ for some invertible but not necessarily known function V ;



(a) An instrumental variable



(b) A nondifferential proxy of the confounder

Figure 1. Example causal diagrams for auxiliary variables.

- (iii) Completeness: for any α , $f(u | x, z; \alpha)$ is complete in z , that is, for any fixed x and square-integrable function g , $E\{g(U) | X = x, Z; \alpha\} = 0$ almost surely if and only if $g(U) = 0$ almost surely.

The exclusion restriction characterizing auxiliary variables is the same condition invoked for the treatment-inducing confounder proxies by Miao, Geng, and Tchetgen Tchetgen (2018) and Tchetgen Tchetgen et al. (2020); it rules out the existence of a direct causal association between the auxiliary variable and the outcome. It is satisfied by instrumental variables and confounder proxies or negative controls. Figure 1 includes two directed acyclic graph (DAG) examples that satisfy the assumption. In Section 3.3 we will illustrate the difference between our use of auxiliary variables and previous proposals.

Equivalence is a high-level assumption stating that the treatment-confounder distribution lies in a model that is identified upon a one-to-one transformation of U . Because ignorability holds conditional on any one-to-one transformation of U , this allows us to use an arbitrary admissible treatment-confounder distribution to identify the treatment effects. The equivalence property restricts the class of treatment-confounder distributions; as one example, it is not met if the dimension of confounders exceeds that of the treatments. Nonetheless, the equivalence property admits a large class of models. In particular, it allows for any factor model or mixture model that is identified, where identification in the context of these models does not imply point identification but rather identification up to a rotation (factor models) or up to label switching (mixture models). Such model assumptions are often used in bioinformatics applications where the unmeasured confounder represents population structure (GWAS) or lab batch effects (Wang et al. 2017; Wang and Blei 2019; Luo and Wei 2019). Identification results for factor and mixture models have been very well established (Anderson and Rubin 1956; Yakowitz and Spragins 1968; Titterton, Smith, and Makov 1985; Kuroki and Pearl 2014). A major limitation of factor models is that they are in general not identified when there are single-treatment confounders or when there are causal relationships among the treatments (Ogburn, Shpitser, and Tchetgen Tchetgen 2019). However, the equivalence assumption can accommodate models that allow for both of these features, for instance, normal mixture models (Yakowitz and Spragins 1968); see the supplementary material for an example.

Completeness is a fundamental concept in statistics (see Lehman and Scheffé 1950), and primitive conditions are readily available in the literature, including the fact that it holds for very general exponential families of distributions and for many

regression models; see for example, Newey and Powell (2003) and D'Haultfœuille (2011). Chen et al. (2014) and Andrews (2017) have shown that if Z and U are continuously distributed and the dimension of Z is larger than that of U , then under a mild regularity condition the completeness condition holds generically in the sense that the set of distributions for which completeness fails has a property analogous to having zero Lebesgue measure. By appealing to such results, completeness holds in a large class of distributions and thus, one may argue that it is commonly satisfied. The role of completeness in this article is analogous to its wide use in a variety of nonparametric and semiparametric identification problems, for instance, in IV regression (Newey and Powell 2003), IV quantile regression (Chernozhukov and Hansen 2005), measurement error problem (Hu and Schennach 2008), missing data (Miao and Tchetgen Tchetgen 2016; D'Haultfœuille 2010), and proximal inference (Miao, Geng, and Tchetgen Tchetgen 2018). Our completeness assumption means that, conditional on X , any variability in U is captured by variability in Z , analogous to the relevance condition in the instrumental variable identification. It is easiest understood in the categorical case. For the binary confounder case, completeness holds if U and Z are correlated within each level of X . When both U and Z have k levels, completeness means that the matrix $[f(u_i | x, z_j)]_{k \times k}$ consisting of the conditional probabilities is invertible. This is stronger than dependence of Z and U given X . Roughly speaking, dependence reveals that variability in U is accompanied by variability in Z , and completeness reinforces that any infinitesimal variability in U is accompanied by variability in Z . As a consequence, completeness in general fails if the number of levels or dimension of Z is smaller than that of U or Z is a coarsening of U . In practice, completeness is more plausible if practitioners measure a rich set of potential auxiliary variables for the purpose of confounding adjustment. In the usual case that the dimension of U is much smaller than that of X , the dimension of Z can also be small. Completeness can be checked in specific models, for instance, in the categorical case. However, Canay, Santos, and Shaikh (2013) show that for unrestricted models the completeness condition is in fact untestable. In the supplementary material, we further elaborate the discussion on completeness and provide both positive and negative examples to facilitate its interpretation and use in practice.

Proposition 1 formalizes the equivalence and completeness conditions for the linear factor model that is widely used in GWAS and computational biology applications.

Proposition 1. Consider a factor model $X = \alpha U + \eta Z + \varepsilon$ for a vector of p observed variables X , q unobserved confounders

U , and $r(\geq q)$ instrumental variables Z such that $Z \perp\!\!\!\perp U \perp\!\!\!\perp \varepsilon$. Without loss of generality we let $E(U) = 0$, $\Sigma_U = I_q$, $E(\varepsilon) = 0$, and Σ_ε be diagonal. Assuming there remain two disjoint submatrices of rank q after deleting any row of α , we have that

- (i) $\alpha\alpha^T$ and Σ_ε are uniquely determined from $\Sigma_{X-\eta Z} = \alpha\alpha^T + \Sigma_\varepsilon$, and any admissible value for α can be written as $\tilde{\alpha} = \alpha R$ with R an arbitrary $q \times q$ orthogonal matrix;
- (ii) if the components of ε are mutually independent and the joint characteristic function of X does not vanish, then any admissible joint distribution can be written as $\tilde{f}(x, u | z) = f(X = x, R^T U = u | Z = z; \alpha)$ with R an arbitrary $q \times q$ orthogonal matrix;
- (iii) if U, Z , and ε^T are normal variables and $\eta^T \gamma$ has full rank of q , then $f(u | x, z) \sim N(\gamma^T x - \gamma^T \eta z, \Sigma)$ and is complete in z , where $\gamma = (\Sigma_{X-\eta Z})^{-1} \alpha$, $\Sigma = I_q - \alpha^T (\Sigma_{X-\eta Z})^{-1} \alpha$.

The first result follows from Anderson and Rubin (1956) by noting that η is identified by the regression of X on Z , the third result can be obtained from the completeness property of exponential families (Newey and Powell 2003, Theorem 2.2), and we prove the second result in the supplementary material. The first two results hold without Z , that is, when $\eta = 0$. This proposition demonstrates that, for the linear factor model, any admissible value for α must be some rotation of the true value and any admissible treatment-confounder distribution must be the joint distribution of X and some rotation of U . Proposition 1 requires that $p \geq 2q + 1$ and that each confounder is correlated with at least three observed variables, and therefore, implies “no single- or dual-treatment confounders.” This is stronger than the “no single-treatment confounder” assumption of Wang and Blei (2019); however, Grimmer, Knox, and Stewart (2020) argue that in fact Wang and Blei (2019) require the much stronger assumption of “no finite-treatment confounding.”

3.2. Identification

Leveraging auxiliary variables gives the following identification result.

Theorem 1. Under Assumptions 1 and 2, for any admissible joint distribution $\tilde{f}(x, u | z)$ that solves $f(x | z) = \int_u \tilde{f}(x, u | z) du$, there exists a unique solution $\tilde{f}(y | u, x)$ to the equation

$$f(y | x, z) = \int_u \tilde{f}(y | u, x) \tilde{f}(u | x, z) du, \tag{4}$$

and the potential outcome distribution is identified by

$$f\{Y(x) = y\} = \int_u \tilde{f}(y | u, x) \tilde{f}(u) du. \tag{5}$$

where $\tilde{f}(u)$ is obtained from $\tilde{f}(x, u | z)$ and $f(z)$.

Although the equivalence and completeness assumptions impose restrictions on the treatment-confounder distribution $f(x, u | z)$, the outcome model $f(y | u, x)$ is left unrestricted in the sense that the parameter space of $f(y | u, x)$ is all possible conditional densities of Y given X and a q -dimensional confounder U . Theorem 1 depicts three steps of the auxiliary variables approach. First we obtain an arbitrary admissible

distribution $\tilde{f}(x, u | z)$; then by solving Equation (4) we identify $\tilde{f}(y | u, x)$, which encodes the treatment effect within each stratum of the confounder; and finally we integrate the stratified effect to obtain the treatment effect in the population. The auxiliary variables approach does not estimate the confounder, or even a surrogate confounder, and thus, dispenses with the need for an infinite number of treatments and avoids the forced positivity violations that were described by D’Amour (2019a, 2019b) and Ogburn, Shpitser, and Tchetgen Tchetgen (2020).

The auxiliary variable is indispensable in the second stage of the approach; without it one has to solve $f(y | x) = \int_u \tilde{f}(y | u, x) \tilde{f}(u | x) du$ for the outcome model. The solution to this equation is not unique given $f(y | x)$ and $\tilde{f}(u | x)$. However, by incorporating an auxiliary variable satisfying the exclusion restriction, we obtain Equation (4), a Fredholm integral equation of the first kind (Kress 1989, chap. 15). The solution of this equation is unique under the completeness condition and thus, identifies the outcome model, up to an invertible transformation of the confounder. Equation (4) also offers testable implications for Assumption 2: if the equation does not have a solution, then Assumption 2 must be partially violated.

Unlike the g -formula (1), we do not identify the true outcome model $f(y | u, x)$ or the true confounder distribution $f(u)$, but instead we obtain $\tilde{f}(y | u, x) = f\{y | V(U) = u, x\}$ and $\tilde{f}(x, u | z) = f\{x, V(U) = u | z\}$ for some invertible transformation $V(U)$. Nonetheless, for any such admissible pair of outcome model and treatment-confounder distribution, we can still identify the potential outcome distribution, because ignorability holds conditional on any such transformation of U . The equivalence assumption guarantees that any admissible distribution $\tilde{f}(x, u | z)$ can be used for identifying the potential outcome distribution; we do not need to use the truth $f(x, u | z)$ and thus, bypass the challenge to identifying it.

Although Theorem 1 shows that the potential outcome distribution is identified, the integral Equation (4) does not admit an analytic solution in general and one has to resort to numerical methods. For instance, Chae, Martin, and Walker (2019) provide an estimation algorithm that is conjectured to provide a consistent estimator of the unknown function under mild conditions. Nonetheless, in certain special cases, a closed-form identification formula can be derived.

Example 1. Suppose p treatments, one confounder, one instrumental variable, and one outcome are generated as $X = \alpha U + \eta Z + \varepsilon$ and $Y = m(X, U, e)$, where (ε^T, U, Z) is a vector of independent normal variables with mean zero, $\Sigma_U = 1$, m is unknown, and $e \perp\!\!\!\perp (\varepsilon^T, U, Z)$. We require that at least three entries of α are nonzero and that $\eta^T \gamma \neq 0$, in which case, the equivalence and completeness assumptions are met according to Proposition 1. Given an admissible value $\tilde{\alpha}$, we let $\tilde{\gamma} = (\Sigma_{X-\eta Z})^{-1} \tilde{\alpha}$, $\tilde{\sigma}^2 = 1 - \tilde{\alpha}^T (\Sigma_{X-\eta Z})^{-1} \tilde{\alpha}$, then $\tilde{f}(u | x, z) \sim N(\tilde{\gamma}^T x - \tilde{\gamma}^T \eta z, \tilde{\sigma}^2)$ is an admissible distribution for $f(u | x, z)$. Let

$$h_1(t) = \int_{-\infty}^{+\infty} \exp(-itz) \phi(z) dz,$$

$$h_2(y, x, t) = -\frac{\tilde{\gamma}^T \eta}{\tilde{\sigma}} \int_{-\infty}^{+\infty} \exp\left\{-it \frac{\tilde{\gamma}^T x - \tilde{\gamma}^T \eta z}{\tilde{\sigma}}\right\} f(y | x, z) dz,$$

be the Fourier transforms of the standard normal density function ϕ and $f(y | x, z)$, respectively, where $i = (-1)^{1/2}$ denotes the imaginary unity. Then the solution to (4) with $\tilde{f}(u | x, z)$ given above is

$$\tilde{f}(y | x, u) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp\left(\frac{itu}{\tilde{\sigma}}\right) \frac{h_2(y, x, t)}{h_1(t)} dt,$$

and the potential outcome distribution is

$$f\{Y(x) = y\} = \int_{-\infty}^{+\infty} \tilde{f}(y | u, x) \phi(u) du.$$

Detailed derivation for [Example 1](#) is deferred to the supplementary material. If a linear outcome model is assumed and the structural causal parameter is of interest, then identification and estimation are simplified as we will discuss in [Section 5.2](#). In the supplementary material, we include an additional example where identification rests on a confounder proxy variable.

3.3. A Comparison to the Conventional Instrumental Variable and Proximal Inference Approaches

We briefly describe the difference between our auxiliary variables approach and the instrumental variable and negative control or proximal inference approaches. In addition to the exclusion restriction ($Z \perp\!\!\!\perp Y | (X, U)$), the instrumental variable approach requires additional assumptions to achieve identification, such as an additive outcome model not allowing for interaction of the treatment and the confounder $E(Y | u, x) = m(x) + u$ as well as completeness in z of $f(x | z)$ (Newey and Powell 2003). Alternative strands of using IV for confounding adjustment include nonseparable outcome models (Imbens and Newey 2009; Chernozhukov and Hansen 2005; Wang and Tchetgen Tchetgen 2018) and local average treatment effect models (Angrist, Imbens, and Rubin 1996; Ogburn, Rotnitzky, and Robins 2015). However, these two approaches typically focus on a single (or binary) treatment and we are not aware of any extensions for multiple treatments. Therefore, we compare our approaches to the additive model, which has a straightforward extension to multiple treatments. The completeness of $f(x | z)$ guarantees uniqueness of the solution to $E(Y | z) = \int_x m(x) f(x | z) dx$, an integral equation identifying $m(x)$. In contrast, our approach does not rest on outcome model restrictions and hence, accommodates interactions. Moreover, the completeness of $f(x | z)$ in z entails at least as many instrumental variables as there are confounded treatments, and requires each confounded treatment to be correlated with at least one instrumental variable. But for our auxiliary variables approach, completeness of $f(u | x, z)$ in z requires the dimension of Z to be as great as that of U , which can be much smaller than that of the treatments. In the special case of a single confounder, completeness of $f(u | x, z)$ can be more plausibly achieved with only one auxiliary variable and with only one treatment correlated with it.

Proximal inference (Miao, Geng, and Tchetgen Tchetgen 2018; Shi et al. 2020; Tchetgen Tchetgen et al. 2020) allows for unrestricted outcome models, but entails at least two confounder proxies (W, Z) with exclusion restrictions: $W \perp\!\!\!\perp (X, Z) | U$ and $Z \perp\!\!\!\perp Y | (X, U)$, even in the single confounder setting.

This approach additionally assumes existence of a function $h(w, y, x)$, called the confounding bridge function, such that $f(y | u, x) = \int_w h(w, y, x) f(w | u) dw$, that is, $h(w, y, x)$ suffices to depict the relationship between the confounding on Y and W . The integral equation $f(y | x, z) = \int_w h(w, y, x) f(w | x, z) dw$ is solved for the confounding bridge $h(w, y, x)$, and the potential outcome distribution is obtained by $f\{Y(x) = y\} = \int_w h(w, y, x) f(w) dw$, where completeness of $f(u | x, z)$ in z is also required for identification of $f\{Y(x)\}$.

A strength of these two approaches is that they leave the treatment-confounder distribution unrestricted. But when the correlation structure of multiple treatments is informative about the presence and nature of confounding, as is generally the case in GWAS and computational biology applications, our method can exploit this correlation structure to remove the confounding bias, while the conventional instrumental variable and proximal inference approaches are agnostic to the treatment-confounder distribution and therefore, unable to leverage any information it contains.

4. Identification Under the Null Treatments Assumption

4.1. Identification

Without auxiliary variables, we let $f(x, u | \alpha)$ denote a model for the treatment-confounder distribution and $f(x; \alpha)$ the resulting marginal distribution indexed by a possibly infinite-dimensional parameter α . Let $\mathcal{C} = \{i : f(u | x) \text{ varies with } x_i\}$ denote the indices of confounded treatments that are associated with the confounder and $\mathcal{A} = \{i : f(y | u, x) \text{ varies with } x_i\}$ the active ones that affect the outcome. A key feature of this identification strategy is that the analyst does not need to know which treatments are confounded or which are active. We make the following assumptions.

- Assumption 3.** (i) Null treatments: the cardinality of the intersection $\mathcal{C} \cap \mathcal{A}$ does not exceed $(|\mathcal{C}| - q)/2$, where $|\mathcal{C}|$ is the cardinality of \mathcal{C} and must be larger than the dimension of U ;
- (ii) Equivalence: for any α , any $\tilde{f}(x, u)$ that solves $f(x; \alpha) = \int_u \tilde{f}(x, u) du$ can be written as $\tilde{f}(x, u) = f\{X = x, V(U) = u; \alpha\}$ for some invertible but not necessarily known function V ;
- (iii) Completeness: for any α , $f(u | x; \alpha)$ is complete in any q -dimensional subvector x_S of the confounded treatments x_C .

Analogous to [Assumption 2](#), the equivalence and completeness assumptions restrict the treatment-confounder distribution, but hold for certain well-known classes of models like factor or mixture models. Under the equivalence assumption, one can identify the confounded treatments set \mathcal{C} by using an arbitrary admissible joint distribution $\tilde{f}(x, u)$ without knowing the truth. The null treatments assumption entails that fewer than half of the confounded treatments can have causal effects on the outcome but does not require knowledge of which treatments are active. The assumption is reasonable in many empirical studies where only a few but not many of the treatments can causally affect the outcome. For example, in GWAS or analyz-

ing electronic health record databases for off-label drugs that improve COVID-19 outcomes, one may expect that most treatments are null without knowing which are active treatments. A counterpart of the null treatments assumption was previously considered by Wang et al. (2017) in the context of effects of a treatment on multiple outcomes. They consider a linear factor model $Y = \alpha U + \beta X + \varepsilon$ for p outcomes (Y) and show that β is identified if only a small proportion of its elements are nonzero. By reversing the roles of treatments and outcome, their approach can be adapted to test but not identify multi-treatment effects, because the coefficient in the regression of treatments on the outcome and the confounder is not the causal effect of interest. Another related concept is the s -sparsity (Kang et al. 2016) used in Mendelian randomization, which assumes that at most s single nucleotide polymorphisms can directly affect the outcome of interest. In contrast, the null treatments assumption does not necessarily require the effects to be sparse and imposes no restrictions on the unconfounded treatments.

Leveraging the null treatments assumption gives the following identification result.

Theorem 2. Under Assumptions 1 and 3, for any joint distribution $\tilde{f}(x, u)$ that solves $f(x) = \int_u \tilde{f}(x, u) du$, there exists a unique solution $\tilde{f}(y | u, x)$ to the equation

$$f(y | x) = \int_u \tilde{f}(y | u, x) \tilde{f}(u | x) du, \tag{6}$$

and the potential outcome distribution is identified by

$$f\{Y(x) = y\} = \int_u \tilde{f}(y | u, x) \tilde{f}(u) du. \tag{7}$$

This theorem states that treatment effects are identified if fewer than half of the confounded treatments can affect the outcome. The outcome model is left unrestricted except for the restriction imposed by the null treatments assumption.

Analogous to the auxiliary variables approaches, the equivalence assumption allows us to use an arbitrary admissible treatment-confounder distribution for identification, the null treatments assumption allows us to construct Fredholm integral equations of the first kind to solve for the outcome model, and the completeness assumption guarantees uniqueness of the solution. Without the null treatments assumption, the solution to (6) is not unique as we noted above.

Theorem 2 holds if we replace q with an integer $s \geq q$ in Assumption 3, in which case, completeness is weakened but the null treatments assumption is strengthened. If the average treatment effect is of interest, we could define the active treatments set as $\mathcal{A} = \{i : E(Y | u, x) \text{ varies with } x_i\}$ and analogously identify $E\{Y(x)\}$. If a linear outcome model is assumed and the structural parameter is of interest, the identification and estimation are simplified. We demonstrate this in Section 5.3.

To illustrate, the following is a simple example where Assumption 3 holds and identification is achieved. We are not aware of any previous identification results for this setting, in particular when no parametric assumption is imposed on the outcome model.

Example 2. Suppose p treatments, one confounder, and one outcome are generated as $X = \alpha U + \varepsilon$ and $Y = m(X, U, e)$,

where (ε^T, U) is a vector of independent normal variables with mean zero, $\Sigma_U = 1$, m is unknown, and $e \perp\!\!\!\perp (\varepsilon, U)$. Suppose the entries of α are nonzero and m can depend on at most $(p - 1)/2$ treatments, but we do not know which ones. In this setting, $f(u | x) \sim N(\gamma^T x, \sigma^2)$ with $\gamma = \Sigma_X^{-1} \alpha$ and $\sigma^2 = 1 - \alpha^T \Sigma_X^{-1} \alpha$. Because the entries of α are nonzero, the equivalence assumption is met (Proposition 1 with $\eta = 0$) and all entries of γ must be nonzero (lemma 2 in the supplementary material); thus, $f(u | x)$ is complete in each treatment. As a result, the potential outcome distributions and treatment effects are identified.

The null treatments approach proceeds by first obtaining an admissible $\tilde{f}(x, u)$, then finding the solution to (6), and finally calculating $f\{Y(x)\}$ from (7). Equation (6) cannot be solved directly as we do not know which treatments are null. As an informal, heuristic intuition for how this identification strategy works in this example, note that, if we knew the identity of any one of the null treatments we could treat it as an auxiliary variable and apply the auxiliary variables approach. In the absence of such knowledge we can imagine using each X as an auxiliary variable; under the null treatments assumption more than half of the X 's must give the same, correct solution to (6). We describe a constructive method that formalizes this idea.

Let x_S denote an arbitrary q -dimensional subvector of x_C and $x_{\bar{S}}$ the rest components of x except for x_S . Given $\tilde{f}(x, u)$ and $f(y | x)$, we solve

$$f(y | x) = \int_u \tilde{f}(y | u, x_{\bar{S}}) \tilde{f}(u | x_S, x_{\bar{S}}) du, \tag{8}$$

for $\tilde{f}(y | u, x_{\bar{S}})$ for all choices of x_S .

Proposition 2. Under Assumptions 1 and 3, for any choice for x_S , if the solution to (8) exists and depends on at most $(|C| - q)/2$ ones of the confounded treatments, then it must solve (6).

4.2. Hypothesis Testing Without Auxiliary Variables and Null Treatments Assumptions

An immediate extension of the null treatments approach provides a test of the sharp null hypothesis of no joint effects, which requires neither auxiliary variables nor the null treatments assumption. The sharp null hypothesis is $\mathbb{H}_0 : f(y | u, x) = f(y | u)$ for all x .

Proposition 3. Under Assumption 1 and (ii)–(iii) of Assumption 3 and given an admissible joint distribution $\tilde{f}(x, u)$, if the null hypothesis \mathbb{H}_0 is correct, then for any q -dimensional subvector x_S of the confounded treatments x_C , the solution to the following equation exists and is unique,

$$f(y | x) = \int_u \tilde{f}(y | u, x_{\bar{S}}) \tilde{f}(u | x_S, x_{\bar{S}}) du, \tag{9}$$

and the solution must satisfy the following equality,

$$f(y) = \int_u \tilde{f}(y | u, x_{\bar{S}}) \tilde{f}(u) du. \tag{10}$$

This result allows us to construct valid hypothesis tests even in the absence of auxiliary variables or a commitment to the

null treatments assumption. Under [Assumption 1](#) and (ii)–(iii) of [Assumption 3](#), evidence against the existence of the solution to (9) or against the equality (10) is evidence against \mathbb{H}_0 . The proof is immediate by noting that, under the null hypothesis \mathbb{H}_0 , the null treatments assumption is trivially satisfied. Thus, if \mathbb{H}_0 is correct, the solution to (9) does not depend x , and the right hand side of (10) identifying the potential outcome distribution $f\{Y(x) = y\}$ must be equal to the observed outcome distribution $f\{Y = y\}$.

5. Estimation

5.1. General Estimation Strategies

In this section we adhere to a common principle of causal inference, and indeed statistics more broadly, that is nicely summed up by Cox and Donnelly (2011) (via Imai and Jiang (2019)):

If an issue can be addressed nonparametrically then it will often be better to tackle it parametrically; however, if it cannot be resolved nonparametrically then it is usually dangerous to resolve it parametrically. (p. 96)

Having proven identification without recourse to parametric outcome models, we will see that estimation does, in fact, require them. This is similar to myriad other causal inference problems, where nonparametric identification is commonly followed by estimation via parametric or sometimes semiparametric models. One key difference is that, in the absence of parametric assumptions, there is no closed-form identifying expression for treatment effects in this setting, because in general there is no closed-form solution to the integral equations that are involved in identification. We first describe an estimation procedure in full generality and then introduce some choices of parametric outcome models that make it feasible in practice.

[Theorem 1](#) points to the following auxiliary variables algorithm for estimation of $f\{Y(x)\}$.

The auxiliary variables algorithm:

- Aux-1* Obtain an arbitrary admissible joint distribution $\tilde{f}(x, u, z)$.
- Aux-2* Use the estimate from Step 1, along with an estimate of $f(y | x, z)$, to solve Equation (4) for an estimate of $\tilde{f}(y | u, x)$.
- Aux-3* Plug the estimate of $\tilde{f}(y | u, x)$ from Step 2 and the estimate of $\tilde{f}(u)$ derived from $\tilde{f}(u, x, z)$ into Equation (5) to estimate $f\{Y(x)\}$.

[Theorem 2](#) points to the null treatments algorithm with a similar set of steps for estimation of $f\{Y(x)\}$.

The null treatments algorithm:

- Null-1* Obtain an arbitrary admissible joint distribution $\tilde{f}(x, u)$.

Null-2 Use the estimate of $\tilde{f}(u | x)$ from Step 1, along with an estimate of $f(y | x)$, to solve Equation (6) for an estimate of $\tilde{f}(y | u, x)$. The constructive method described in [Proposition 2](#) can be implemented to solve (6).

Null-3 Plug the estimate of $\tilde{f}(u)$ from Step 1 and $\tilde{f}(y | u, x)$ from Step 2 into Equation (7) to estimate $f\{Y(x)\}$.

Steps *Aux-1* and *Null-1* require only the equivalence assumption, which places nontrivial restrictions on the treatment-confounder distribution. To estimate $\tilde{f}(x, u)$, one needs to correctly specify a treatment-confounder model that meets the equivalence assumption, such as a factor or mixture model. Under standard factor or mixture models, estimation of $\tilde{f}(x, u)$ is well established, and we refer to the large existing body of literature for estimation techniques and properties (Anderson and Rubin 1956; Kim and Mueller 1978; Titterton, Smith, and Makov 1985). Note that, crucially, Step 1 estimates the distribution of U (joint with X), but does not estimate U itself. This is advantageous as it does not require infinite number of treatments and engender the resulting positivity problems.

Steps *Aux-2* and *Null-2* involve, first, estimating $f(y | x)$ or $f(y | x, z)$, which can be done parametrically or nonparametrically using standard density estimation techniques. More challenging is the second step, solving integral Equations (4) and (6), which do not admit analytic solutions in general. These equations are of the form of Fredholm integral equations of the first kind (Kress 1989, chap. 15). This kind of equation is known to be ill-posed due to noncontinuity of the solution and difficulty of computation. In the contexts of nonparametric instrumental variable regression, regularization methods have been established to solve the equation and we refer to Newey and Powell (2003); Carrasco, Florens, and Renault (2007) for a broad view of this problem. Numerical solution to such equations is an active area of mathematical and statistical research and is largely beyond the scope of this article. However, we note that Chae, Martin, and Walker (2019) provide R code for a numerical method that is conjectured to provide a consistent estimator of the unknown function under mild conditions. In the next sections, we describe modeling assumptions under which the integral equation can be avoided altogether.

Steps *Aux-3* and *Null-3* are essentially applications of the g -formula; estimation of this integral is standard in causal inference problems.

Below we provide two examples of how parametric models—in this case linear—can obviate the need to solve integral equations, permit estimation using standard software, and admit consistent estimators.

5.2. The Auxiliary Variables Approach with Linear Models

Consider the following model for a p -dimensional treatment X , a q -dimensional confounder U , and an r -dimensional instrumental variable Z , with $p \geq 2q + 1$ and $r \geq q$:

$$\begin{aligned} X &= \alpha U + \eta Z + \varepsilon, & \Sigma_U &= I_q, & E(U) &= 0, \\ U &\perp\!\!\!\perp Z \perp\!\!\!\perp \varepsilon, & \Sigma_\varepsilon &\text{diagonal}, & & \end{aligned} \quad (11)$$

there remain two disjoint submatrices of rank q

after deleting any row of α , (12)

$\eta^T \gamma$ has full rank of q , where

$$\gamma = (\Sigma_{X-\eta Z})^{-1} \alpha = (\Sigma_\varepsilon + \alpha \alpha^T)^{-1} \alpha, \quad (13)$$

$$E(Y | U, X, Z) = \beta^T X + \delta^T U. \quad (14)$$

This is the IV setting from [Example 1](#). Intercepts are not included in the models as one can center (X, Y, Z) to have mean zero. If (ε^T, U, Z) are normally distributed, then (11)–(12) imply equivalence and (13) implies completeness in [Theorem 1](#), and as a special case, identification and estimation of $f\{Y(x)\}$ follows from [Theorem 1](#) and the auxiliary variables algorithm, respectively. However, the estimation procedure below works even when the error distributions are left unspecified, in which case (11)–(13) do not suffice for identification of $f\{Y(x)\}$. We additionally assume the linear outcome model (14) and focus on the structural parameter β but not the potential outcome distribution $f\{Y(x)\}$. Then (11)–(13), viewed as a parallel version of [Assumption 2](#), guarantee identification of β . Condition (13) in principle can be tested after obtaining an estimator of η and γ . Note that, r may be smaller than p , in which case the conventional IV estimator $\hat{\beta}_{iv} = (Z^T X)^{-1} Z^T Y$ does not work.

Estimation of β is parallel to the auxiliary variables algorithm. We first obtain $\hat{\eta}$ by regression of X on Z and obtain $\hat{\gamma}$ by factor analysis of the residuals $X - \hat{\eta}Z$. There must exist some orthogonal matrix R so that $\hat{\gamma}$ converges to γR . This corresponds to *Aux-1*. Let $(\hat{\xi}^X, \hat{\xi}^Z)$ denote the coefficients by regression of Y on (X, Z) and $(\hat{\xi}^X, \hat{\xi}^Z)$ be the corresponding estimator. Note that $\hat{\xi}^X = \beta + \gamma \delta$ and $\hat{\xi}^Z = -\eta^T \gamma \delta$, therefore, we solve

$$\begin{aligned} \hat{\xi}^X &= \hat{\beta} + \hat{\gamma} \hat{\delta}, \\ \hat{\xi}^Z &= -\hat{\eta}^T \hat{\gamma} \hat{\delta} \end{aligned} \quad (15)$$

for $(\hat{\beta}, \hat{\delta})$. This corresponds to *Aux-2*: estimation of $f(y | x, z)$ is replaced by a linear regression of Y on (X, Z) and solving the integral equation for $\tilde{f}(y | u, x)$ is replaced by solving linear equations for finite-dimensional parameters (β, δ) . We finally obtain

$$\hat{\beta} = \hat{\xi}^X + \hat{\gamma} (\hat{\gamma}^T \hat{\eta} \hat{\eta}^T \hat{\gamma})^{-1} \hat{\gamma}^T \hat{\eta} \hat{\xi}^Z.$$

In the special case that the dimension of the instrumental variable Z equals that of the confounder U , we obtain $\hat{\beta} = \hat{\xi}^X + \hat{\gamma} (\hat{\eta}^T \hat{\gamma})^{-1} \hat{\xi}^Z$. Consistency and asymptotic normality follows from that of $(\hat{\xi}^X, \hat{\xi}^Z, \hat{\gamma}, \hat{\eta})$. Routine R software such as `factanal` and `lm` can be implemented for factor analysis and linear regression, respectively, and the variance of estimators can be bootstrapped.

5.3. The Null Treatments Approach with Linear Models

Consider the linear models:

$$\begin{aligned} X &= \alpha U + \varepsilon, & E(Y | X, U) &= \beta^T X + \delta^T U, \\ \Sigma_U &= I_q, & E(U) &= 0, & U &\perp \varepsilon, & \Sigma_\varepsilon &\text{diagonal,} \end{aligned} \quad (16)$$

where U and ε are not necessarily normally distributed. The coefficient β encoding the average treatment effects is of interest.

We let $\gamma = \Sigma_X^{-1} \alpha$, which denotes the coefficients of regressing U on X . Hereafter, we use A_i to denote the i th row of a matrix or a column vector A . Let $\mathcal{C} = \{i : \alpha_i \text{ is not a zero vector}\}$ denote the indices of confounded treatments, $|\mathcal{C}|$ the cardinality of \mathcal{C} , and $\gamma_{\mathcal{C}}$ the submatrix consisting of the corresponding rows of γ . We have the following result.

Theorem 3. The parameter β is identified under model (16) and the following assumptions:

- (i) at most $(|\mathcal{C}| - q)/2$ entries of $\beta_{\mathcal{C}}$ are nonzero;
- (ii) after deleting any row, there remain two disjoint submatrices of α of full rank;
- (iii) any submatrix of $\gamma_{\mathcal{C}}$ consisting of q rows has full rank.

The identification result is not compromised if q is replaced with $s \geq q$ in the conditions. If the errors are normally distributed, then conditions (i)–(iii) of [Theorem 3](#) imply the null treatments, the equivalence and the completeness conditions in [Assumption 3](#), respectively, and identification and estimation of $f\{Y(x)\}$ follows from [Theorem 2](#) and the null treatments algorithm, respectively. The estimation procedure below works for unspecified error distributions, in which case conditions (i)–(iii) in [Theorem 3](#) no longer suffice for identification of $f\{Y(x)\}$, but as a parallel version of [Assumption 3](#) they guarantee identification of β . Condition (iii) of [Theorem 3](#) in principle can be tested after obtaining an estimator of γ .

Estimation of β is parallel to the null treatments algorithm. Let ξ denote the coefficients by regression of Y on X . Given n independent and identically distributed samples, we first obtain $\hat{\gamma}$ by factor analysis of X ; this corresponds to *Null-1*. Then we obtain $\hat{\xi}$ by regression of Y on X , which can be viewed as a crude estimator of β with asymptotic bias $\gamma \delta$. Note that $\xi_{\mathcal{C}} = \beta_{\mathcal{C}} + \gamma_{\mathcal{C}} \delta$ for confounded treatments, then under the null treatments assumption (i) of [Theorem 3](#), estimation of $\hat{\delta}$ can be cast as a standard robust linear regression (Rousseeuw and Leroy 2005) given $\hat{\xi}$ and $\hat{\gamma}$: $\hat{\gamma}_{\mathcal{C}}$ is the design matrix and $\hat{\xi}_{\mathcal{C}}$ is observations with outliers corresponding to nonzero entries of $\beta_{\mathcal{C}}$. This corresponds to *Null-2*, where the complexity of solving integral equations is resolved by robust linear regression. Specifically, given a $n^{1/2}$ -consistent estimator $(\hat{\xi}, \hat{\gamma})$, we solve

$$\begin{aligned} \hat{\delta}^{\text{lms}} &= \arg \min_{\hat{\delta}} \text{median} \{(\hat{\xi}_i - \hat{\gamma}_i \delta)^2, i \in \hat{\mathcal{C}}\}, \\ \hat{\mathcal{C}} &= \{i : \|\hat{\gamma}_i\|_2^2 > \log(n)/n\}, \end{aligned} \quad (17)$$

which is a least median of squares estimator minimizing median of the squared errors $(\hat{\xi}_i - \hat{\gamma}_i \delta)^2$ among the confounded treatments consistently selected by $\hat{\mathcal{C}}$. For the robust linear regression, the least quantile of squares, least trimmed squares, and the S-estimator can also be used to solve for δ , which we refer to Rousseeuw and Leroy (2005, chap. 3.4). The corresponding estimate of β is $\hat{\beta}^{\text{lms}} = \hat{\xi} - \hat{\gamma} \hat{\delta}^{\text{lms}}$. In the supplementary material, we show consistency of $(\hat{\delta}^{\text{lms}}, \hat{\beta}^{\text{lms}})$ under the assumptions of [Theorem 3](#) and an additional regularity condition given $n^{1/2}$ -consistency of $(\hat{\xi}, \hat{\gamma})$. However, they are not necessarily asymptotically normal as we observe in numerical experiments, even if $(\hat{\xi}, \hat{\gamma})$ are. Therefore, to promote asymptotic normality we use $\hat{\beta}^{\text{lms}}$ as an initial value to select the null treatments and to update estimates of δ and β as follows: obtain $\hat{\delta}$ by solving

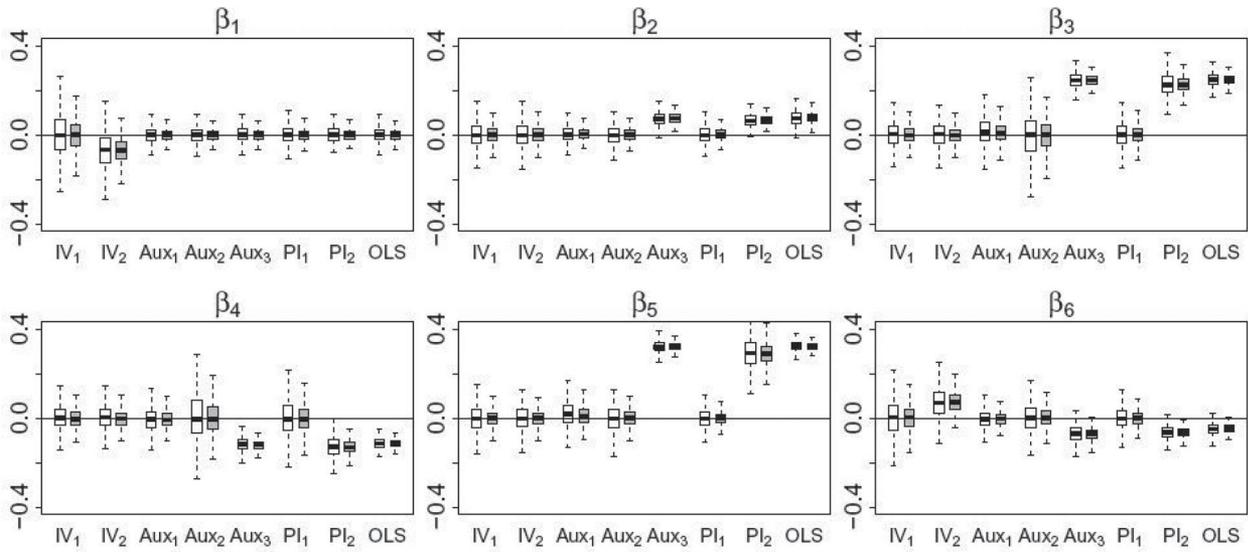


Figure 2. Bias of estimators when the exclusion restriction holds. White boxes are for sample size 1000 and gray ones for 2000.

$\hat{\xi}_i = \hat{\gamma}_i \hat{\delta}$ (by OLS) that corresponds to the smallest $(|\hat{C}| + q)/2$ entries of $|\hat{\beta}^{\text{lms}}|$ and obtain the ultimate estimator $\hat{\beta} = \hat{\xi} - \hat{\gamma} \hat{\delta}$. Routine R software `factanal` and `lqs` can be implemented for factor analysis and robust linear regression, respectively, and the variance of the ultimate estimator can be bootstrapped.

6. Simulations

6.1. The Auxiliary Variables Setting

We evaluate performance of the proposed methods via simulations. For the auxiliary variables setting, two confounders (U), six instrumental variables (Z), six treatments (X), an outcome (Y), and two outcome-inducing confounder proxies (W) are generated as follows,

$$\begin{aligned} X &= \alpha U + \eta Z + \varepsilon_X, & Y &= \beta^T X + \delta_Y U + \varepsilon_Y, \\ W &= \delta_W U + \varepsilon_W, \end{aligned} \quad (18)$$

$$\begin{aligned} U &\sim N(0, I_2), & Z &\sim N(0, I_6), & \varepsilon_X &\sim N(0, I_6), \\ \varepsilon_Y &\sim N(0, 1); & \varepsilon_W &\sim N(0, I_2); \end{aligned} \quad (19)$$

$$\alpha = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 1.5 & 1 \\ 2 & -2 \\ 2.5 & 1 \\ 2 & -1 \end{pmatrix}, \quad \eta = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

$$\beta = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad \delta_Y = (1, 1), \quad \delta_W = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}.$$

Under this setting, (X_2, \dots, X_6) are confounded but X_1 is not. For estimation, we consider eight methods:

- IV₁ conventional IV approach using all six IVs;
- IV₂ conventional IV approach using five IVs (Z_1, \dots, Z_5) and treating (X_6, Z_6) as covariates;

- Aux₁ the proposed auxiliary variables approach assuming two factors and using all six IVs;
- Aux₂ the auxiliary variables approach assuming two factors, using (Z_5, Z_6) as IVs and (Z_1, \dots, Z_4) as covariates;
- Aux₃ the auxiliary variables approach with one factor, using Z_6 as an IV and (Z_1, \dots, Z_5) as covariates;
- PI₁ proximal inference using (Z_5, Z_6) and (W_1, W_2) as the treatment- and outcome-inducing confounder proxies, respectively, and (Z_1, \dots, Z_4) as covariates;
- PI₂ proximal inference using Z_6 and W_1 as the treatment- and outcome-inducing confounder proxy, respectively, and (Z_1, \dots, Z_5) as covariates;
- OLS ordinary least squares estimation by regression Y on X and Z .

Two-stage least squares are used in the above IV or PI methods. Because Grimmer, Knox, and Stewart (2020) have shown that the deconfounder (Wang and Blei 2019) is asymptotically equivalent to and does not outperform OLS, we do not include a separate comparison with the deconfounder. We replicate 1000 simulations at sample size 1000 and 2000. Figure 2 summarizes bias of the estimators of each parameter. As expected, IV₁, Aux₁, Aux₂, and PI₁ perform well with little bias for estimation of all parameters, because sufficient number of IVs or confounder proxies are used in these four methods and the number of factors are correctly specified in Aux₁ and Aux₂. Note that, Aux₂ uses only two IVs while IV₁ uses all six IVs and PI₁ uses two additional confounder proxies. We also compute the bootstrap confidence interval and evaluate the coverage probability for Aux₁ and Aux₂, summarized in Table 1. The 95% bootstrap confidence interval has coverage probabilities close to the nominal level of 0.95 under a moderate sample size. When using the same set of IVs, we might expect Aux₁ to be more efficient than IV₁ as the former additionally incorporates the internal dependence structure of the treatments.¹ However, in our simulations when the sample size is increased to 10,000, we observe that Aux₁ has a larger mean squared error than IV₁ for estimation of

¹This is conjectured by an anonymous reviewer.

Table 1. Coverage probability of the 95% bootstrap confidence interval based on Aux₁ and Aux₂

	β_1	β_2	β_3	β_4	β_5	β_6
Aux ₁	0.948	0.940	0.940	0.970	0.933	0.958
	0.943	0.949	0.954	0.950	0.933	0.949
Aux ₂	0.946	0.966	0.970	0.963	0.972	0.960
	0.942	0.959	0.958	0.953	0.950	0.942

Note: For each estimator, the first row is for sample size 1000 the second for 2000.

(β_3, β_5) and Aux₁ has a larger mean squared error than Aux₂ for estimation of β_5 . It can also be seen from the boxplots of the bias in Figure 2. This may be because the proposed auxiliary variables estimation is not efficient. Therefore, in the future it is of great interest to theoretically establish the semiparametric efficiency bound and the efficient estimator under the auxiliary variables assumption and compare to other competing methods.

In contrast, IV₂ fails to consistently estimate β_6 because the corresponding IV (Z_6) is not correctly used but treated as a covariate, but surprisingly, IV₂ is also biased for estimation of β_1 that is not confounded and can be estimated very well by all the other methods. Likewise, PI₂ is biased for estimation of (β_2, \dots, β_5) due to insufficient number of confounder proxies. Nonetheless, PI₂ has little bias for estimation of β_6 . This is because Z_6 used in PI₂ is a valid IV for X_6 and the PI method is consistent even if the confounder proxies are inadequate, a property previously shown by Miao, Shi, and Tchetgen Tchetgen (2018). Except for β_1 , Aux₃ is biased for estimation of the confounded parameters because the number of factors and IVs are incorrectly specified. As expected, OLS is biased for estimation of all parameters except for β_1 .

We also evaluate performance of the estimators when the IVs have a direct effect on the outcome. In this case, the outcome model is changed to $Y = \beta^T X + \lambda Z + \delta_Y U + \varepsilon_Y$ with $\lambda = (0.2, \dots, 0.2, 0.3)$, while the other settings of the data generating process remain the same. Figure S.1 in the supplementary material summarizes bias of the estimators. All estimators are biased for estimation of the confounded treatment effects, and no estimator outperforms the others in terms of bias. However, the IV estimators are also biased for estimation of β_1 while the other estimators have little bias.

In summary, all of conventional IV, proximal inference, and the proposed auxiliary variables approaches rely on the exclusion restriction assumption. If the exclusion restriction and the required conditions are satisfied, each of these approaches can properly address the multi-treatment confounding. The conventional IV entails at least as many IVs as treatments, otherwise it can be biased even for unconfounded treatments; proximal inference needs at least as many treatment- and outcome-inducing proxies as confounders; the proposed auxiliary variables approach does not rest on outcome-inducing confounder proxies but rests on a factor model and correct specification of the factor number. Therefore, to obtain a reliable causal conclusion in practice, we recommend implementing and comparing multiple applicable methods.

6.2. The Null Treatments Setting

We generate 2 confounders (U), 8 treatments (X), and an outcome (Y) as follows,

$$\begin{aligned}
 X &= \alpha U + \varepsilon_X, & Y &= \beta^T X + \delta_Y U + \varepsilon_Y, \\
 U &\sim N(0, I_2), & \varepsilon_X &\sim N(0, I_6), & \varepsilon_Y &\sim N(0, 1); \\
 \alpha^T &= \begin{pmatrix} 0 & 0.4 & 0.8 & 1.2 & 1.5 & -0.4 & -0.8 & -1.2 \\ 0 & 0.2 & 0.4 & 0.6 & 0.8 & -0.5 & -1.0 & -1.2 \end{pmatrix}, & \delta_Y &= (1, 1).
 \end{aligned}$$

We consider two choices for β ,

Case 1: $\beta_1 = \beta_2 = \beta_3 = 1, \beta_4 = \dots = \beta_8 = 0$;

Case 2: $\beta_1 = \beta_2 = \beta_3 = 1, \beta_4 = \beta_5 = 0.2, \beta_6 = \beta_7 = \beta_8 = 0$.

Under these settings, (X_2, \dots, X_8) are confounded but X_1 is not; the null treatments assumption is satisfied in Case 1 as only two of the confounded treatments are active, but it is violated in Case 2 where β_4 and β_5 also have a small effect on Y .

For estimation, three methods are used: the null treatments estimation with correct dimension of the confounder (Null₁), the null treatments estimation with one confounder (Null₂), and OLS. Because no auxiliary variables are generated, we do not compare to IV, the auxiliary variables, or proximal inference approaches. We replicate 1000 simulations at sample size 2000 and 5000. Figures 3 and Figure S.2 (in the supplementary material) summarize bias of the estimators for Case 1 and Case 2, respectively. As expected, for estimation of β_1 that is not confounded, both OLS and the null treatments approach have little bias in both cases even if the number of confounders is not correctly specified. In Case 1 where the null treatments assumption is met, Null₁ has little bias because the number of confounders is correctly specified, and as shown in Table 2, the 95% bootstrap confidence interval based on Null₁ has coverage probabilities approximate to the nominal level of 0.95. But Null₂ is biased because the dimension of the confounder is specified to be smaller than the truth. The bias is smaller than the OLS, although this is not theoretically guaranteed. If the dimension of the confounder is larger than the truth, the factor analysis fails and so does the null treatments estimation. In Case 2, the null treatments assumption is violated because more than half of the confounded treatments are active. In this case, the null treatments estimation is in general biased for the confounded treatments; the bias could be larger or smaller than OLS. Therefore, to apply the null treatments estimation, one has to first assure that the majority of the treatments are null or very close to zero. Both the auxiliary variables and the null treatments approaches rest on correct specification of number of unmeasured confounders. If it is not known with high confidence, we refer to Bai and Ng (2002); Chen, Li, and Fu (2012); Owen et al. (2016) and McLachlan, Lee, and Rathnayake (2019) for the estimation methods, and we recommend assessing robustness of estimation by varying the specification in data analyses.

7. Application to a Mouse Obesity Study

For further illustration, we reanalyze a mouse obesity dataset described by Wang et al. (2006), where the effect of gene expressions on the body weight of F2 mice is of interest. Unmeasured confounding may arise in such gene expression studies due to batch effects or unmeasured phenotypes. The data we use are collected from 227 mice, including the body weight (Y), 17 gene expressions (X), and five single nucleotide polymorphisms (Z); see the supplementary material for a complete list of these

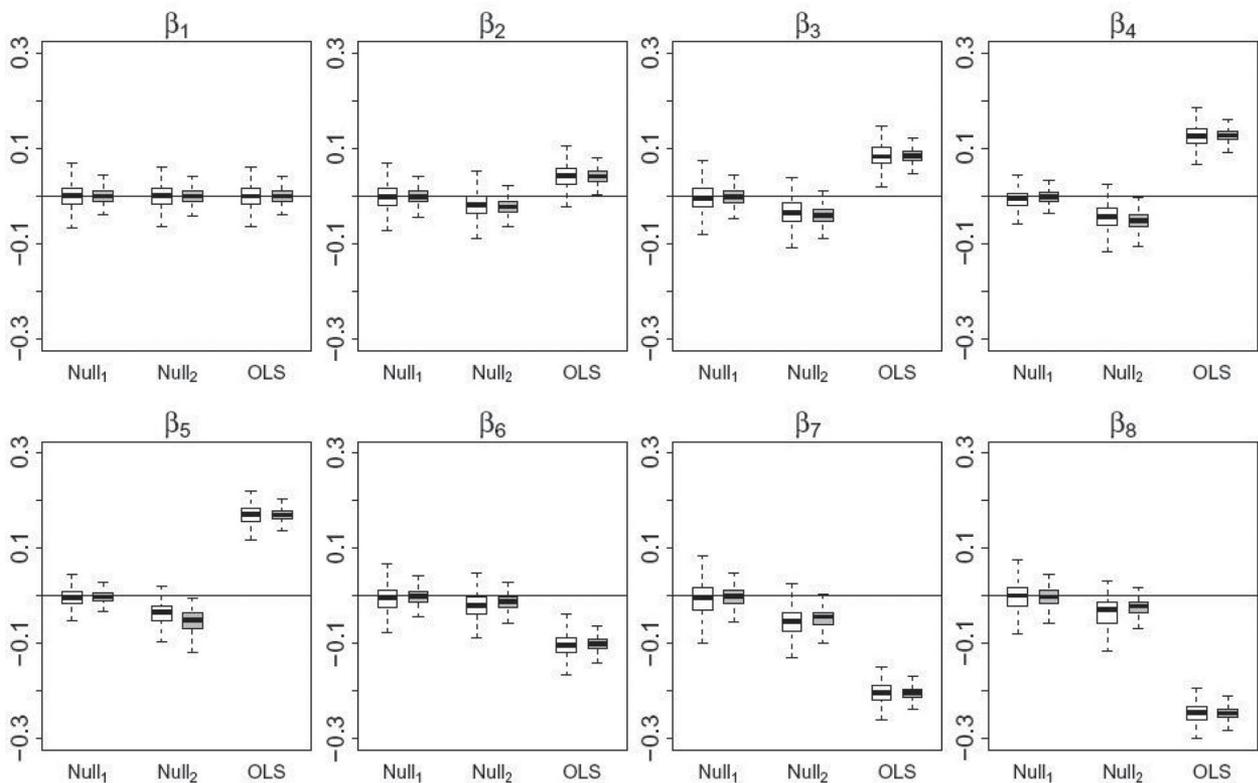


Figure 3. Bias of estimators in Case 1. White boxes are for sample size 2000 and gray ones for 5000.

Table 2. Coverage probability of the 95% bootstrap confidence interval for Case 1

	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
Null ₁	0.967	0.970	0.973	0.975	0.987	0.974	0.990	0.983
	0.943	0.958	0.959	0.967	0.969	0.961	0.963	0.975

Note: The first row is for sample size 2000 and the second for 5000.

variables. As previously selected by Lin, Feng, and Li (2015), the 17 genes are likely to affect mouse weight and the five single nucleotide polymorphisms are potential instrumental variables. We further evaluate the effects of these genes on mouse weight by adopting a factor model that is widely used to characterize the unmeasured confounding in gene expression studies (Gagnon-Bartsch and Speed 2012; Wang et al. 2017). We assume a linear outcome model and estimate the parameters with three methods: the auxiliary variables approach using single nucleotide polymorphisms as instrumental variables, the null treatments approach assuming that fewer than half of genes can affect the mouse weight, and ordinary least squares. We also compute the bootstrap confidence intervals.

Figure 4 presents the point estimates and their significance for the 17 genes, when the factor number is specified as one for the auxiliary variables and the null treatments estimation. Detailed results including point and interval estimates are relegated to the supplementary material. All three methods agree with positive and significant effects of *Gstm2*, *Sirpa*, and *2010002N04Rik*, and a negative effect of *Dscam*. Additionally, the auxiliary variables estimation also indicates negative and significant effects of *Igf1bp2*, *Avpr1a*, *Abca8a*, and *Irx3*; the null treatments estimation suggests a potentially

positive effect of *Gpld1*. These results reinforce previous findings about the effects of genes on obesity. For instance, recent studies show that *Igf1bp2* (Insulin-like growth factor binding protein 2) protects against the development of obesity (Wheatcroft et al. 2007); *Gpld1* (Phosphatidylinositol-glycan-specific phospholipase D) is associated with the change in insulin sensitivity in response to the low-fat diet (Gray et al. 2008); and *Irx3* (Iroquois homeobox gene 3) is associated with lifestyle changes and plays a crucial role in determining weight via energy balance (Schneeberger 2019).

However, the significance test (Kim and Mueller 1978, chap. IV) of the hypothesis that one factor is sufficient is rejected, indicating that either the number of factors is too small or there exists model misspecification. Figure S.3 in the supplementary material shows the results when the number of factors is increased to two and three. With two factors, *Gstm2*, *2010002N04Rik*, *Igf1bp2*, and *Avpr1a* remain significant in the auxiliary variables analysis, and *Gstm2* and *Dscam* remain significant in the null treatments analysis. With three factors, all estimates in both the auxiliary variables and the null treatments analyses are no longer significant. In summary, the association between the 17 genes and mice obesity can be explained by three or more unmeasured confounders. But if there exist only one or two confounders, *Gstm2*, *Sirpa*, *2010002N04Rik*, *Dscam*, *Igf1bp2*, *Avpr1a*, *Abca8a*, *Irx3*, and *Gpld1* have a potential causal association with mouse obesity, which cannot be completely attributed to confounding. In future studies, it is of interest to identify the potential confounders and to use additional data to more confidently estimate effects of these nine genes.

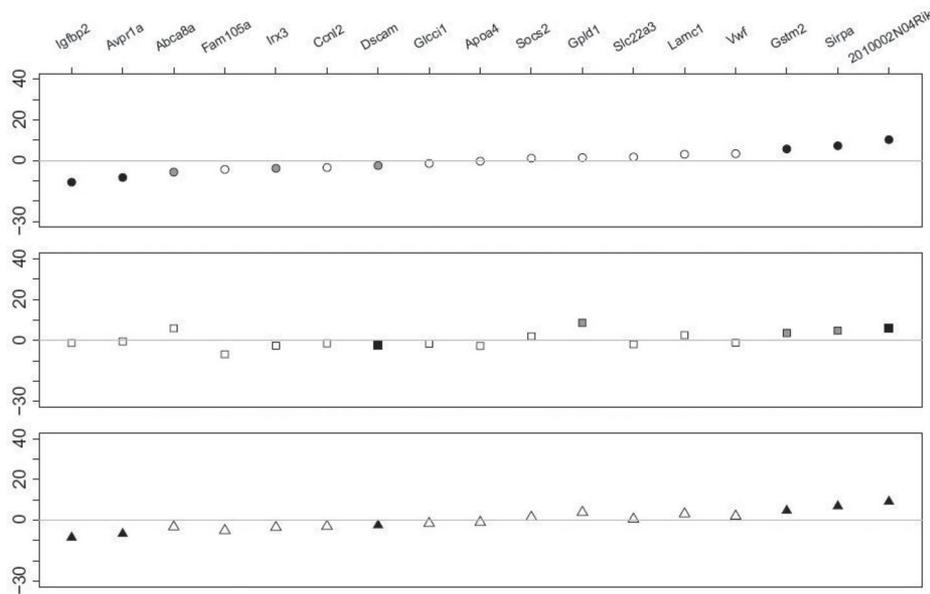


Figure 4. Effect estimates for 17 genes when one factor is used in analyses. The first panel is for the auxiliary variables estimation, the second for the null treatments estimation, and the last for ordinary least squares estimation. Black points are for significant estimates at level of 0.05, gray ones for 0.1, and white ones for estimates not significant at 0.1.

8. Discussion

In this article, we extend results that had previously been developed for the identification of treatment effects in the presence of unmeasured confounding in the single-treatment to the multi-treatment setting, and we extend the parametric approach of Wang et al. (2017) to identification of multi-treatment effects with an unrestricted outcome model. We demonstrate a novel framework using integral equations and completeness conditions to establish identification in causal inference problems.

We have assumed that the number of confounders is known, which is realistic in confounder measurement error or misclassification problems. When it is not known a priori, consistent estimation of the number of confounders has been well established by Bai and Ng (2002) under factor models and by Chen, Li, and Fu (2012) under mixture models. The R software `factanal` provides a significance test of whether the number of factors in a factor model is sufficient to capture the full dimensionality of the dataset. We also refer to Owen et al. (2016) and McLachlan, Lee, and Rathnayake (2019) for a comprehensive literature review related to choosing the number of confounders in practice. We also recommend conducting a sensitivity analysis like in the simulations and the application by altering the specification of the number of confounders to assess the robustness of the substantive conclusion.

Our identification framework rests on the auxiliary variables or the null treatments assumption. These assumptions are partially testable: a heuristic approach, taking Equation (4) as an example, is to check whether a solution exists. This can be achieved by obtaining a solution $\tilde{f}(y | u, x)$ that minimizes the mean squared error $\|f(y | x, z) - \int_u \tilde{f}(y | u, x) \tilde{f}(u | x, z) du\|^2$ and checking how far away the error is from zero to assess how well the solution fits the equation. This is a typical goodness-of-fit test if all models are parametric. However, in nonparametric models statistical inference for the integral equation is quite

difficult and we leave the development of falsification tests for future research. Even if both the auxiliary variables and the null treatments assumptions fail to hold, we describe how to test whether treatment effects exist. The proposed estimation strategies can also be used to test whether unmeasured confounding is present, by assessing how far the proposed estimates are from the crude ones.

Our identification results lead to feasible estimation methods under parametric estimation assumptions. The proposed estimation methods, comprised of standard factor analysis, linear and robust linear regression, inherit properties from the classical theory of statistical inference; these methods work well in simulations and an application. However, statistical inference for nonparametric and semiparametric models remains to be studied. We have considered fixed dimensions of treatments and confounders, and extensions to large and high-dimensional settings are of both theoretical and practical interest.

Supplementary Material

Supplementary material online includes proof of theorems and propositions, useful lemmas, discussion and examples on the completeness condition, consistency of the least median of squares estimator, discussion on identification of a parametric model for a binary outcome, details for examples, additional results for simulations and the application, and codes for reproducing the results in this article.

Acknowledgments

We are grateful for comments from the editors and three anonymous reviewers. We thank Eric Tchetgen Tchetgen for comments on early versions of this article.

Funding

This work was partially supported by Beijing Natural Science Foundation (Z190001) and National Natural Science Foundation of China (12071015, 81773546), and ONR grants N00014-18-1-2760 and N00014-21-1-2820.

ORCID

Orcid Wang Miao  <http://orcid.org/0000-0002-6221-2717>
 Elizabeth Ogburn  <http://orcid.org/0000-0001-7036-7973>
 Xiao-Hua Zhou  <http://orcid.org/0000-0001-7935-1222>

References

- Alter, O., Brown, P. O., and Botstein, D. (2000), “Singular Value Decomposition for Genome-Wide Expression Data Processing and Modeling,” *Proceedings of the National Academy of Sciences*, 97, 10101–10106. [2]
- Anderson, T. W., and Rubin, H. (1956), “Statistical Inference in Factor Analysis,” in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 5: Contributions to Econometrics, Industrial Research, and Psychometry*, pp. 111–150. Berkeley, CA: University of California Press. [4,5,8]
- Andrews, D. W. (2017), “Examples of L^2 -Complete and Noundedly-Complete Distributions,” *Journal of Econometrics*, 199, 213–220. [4]
- Angrist, J., Imbens, G., and Rubin, D. (1996), “Identification of Causal Effects Using Instrumental Variables,” *Journal of the American Statistical Association*, 91, 444–455. [1,6]
- Athey, S., Imbens, G. W., and Pollmann, M. (2019), “Comment on: ‘The blessings of multiple causes’ by Yixin Wang and David M. Blei,” *Journal of the American Statistical Association*, 114, 1602–1604. [2]
- Bai, J. and Ng, S. (2002), “Determining the Number of Factors in Approximate Factor Models,” *Econometrica*, 70, 191–221. [11,13]
- Balke, A., and Pearl, J. (1997), “Bounds on Treatment Effects from Studies with Imperfect Compliance,” *Journal of the American Statistical Association*, 92, 1171–1176. [1]
- Canay, I. A., Santos, A., and Shaikh, A. M. (2013), “On the Testability of Identification in Some Nonparametric Models with Endogeneity,” *Econometrica*, 81, 2535–2559. [4]
- Carrasco, M., Florens, J. P., and Renault, E. (2007), “Linear Inverse Problems in Structural Econometrics Estimation Based on Spectral Decomposition and Regularization,” in *Handbook of Econometrics*, eds. J. J. Heckman and E. Leamer (Vol. 6B, pp. 5633–5751). Amsterdam: Elsevier. [8]
- Čevid, D., Bühlmann, P., and Meinshausen, N. (2020), “Spectral Deconfounding via Perturbed Sparse Linear Models,” *Journal of Machine Learning Research*, 21, 1–41. [2]
- Chae, M., Martin, R., and Walker, S. G. (2019), “On an Algorithm for Solving Fredholm Integrals of the First Kind,” *Statistics and Computing*, 29, 645–654. [5,8]
- Chen, J., Li, P., and Fu, Y. (2012), “Inference on the Order of a Normal Mixture,” *Journal of the American Statistical Association*, 107, 1096–1105. [11,13]
- Chen, X., Chernozhukov, V., Lee, S., and Newey, W. K. (2014), “Local Identification of Nonparametric and Semiparametric Models,” *Econometrica*, 82, 785–809. [4]
- Chernozhukov, V., and Hansen, C. (2005), “An IV Model of Quantile Treatment Effects,” *Econometrica*, 73, 245–261. [4,6]
- Chernozhukov, V., Hansen, C., Liao, Y. (2017), “A Lava Attack on the Recovery of Sums of Dense and Sparse Signals,” *The Annals of Statistics*, 45, 39–76. [2]
- Cornfield, J., Haenszel, W., Hammond, E. C., Lilienfeld, A. M., Shimkin, M. B., and Wynder, E. L. (1959), “Smoking and Lung Cancer: Recent Evidence and a Discussion of Some Questions,” *Journal of the National Cancer Institute*, 22, 173–203. [1]
- Cox, D. R., and Donnelly, C. A. (2011), *Principles of Applied Statistics*, Cambridge: Cambridge University Press. [8]
- D’Amour, A. (2019a), “On Multi-Cause Causal Inference with Unobserved Confounding: Counterexamples, Impossibility, and Alternatives,” in *The 22nd International Conference on Artificial Intelligence and Statistics*, pp. 3478–3486. Okinawa, Japan. [2,3,5]
- D’Amour, A. (2019b), “Comment: Reflections on the Deconfounder,” *Journal of the American Statistical Association*, 114, 1597–1601. [2,5]
- D’Haultfoeulle, X. (2010), “A New Instrumental Method for Dealing with Endogenous Selection,” *Journal of Econometrics*, 154, 1–15. [4]
- “On the Completeness Condition in Nonparametric Instrumental Problems,” *Econometric Theory*, 27, 460–471. [4]
- Didelez, V., and Sheehan, N. (2007), “Mendelian Randomization as an Instrumental Variable Approach to Causal Inference,” *Statistical Methods in Medical Research*, 16, 309–330. [1]
- Ding, P., and Vanderweele, T. J. (2014), “Generalized Cornfield Conditions for the Risk Difference,” *Biometrika*, 101, 971–977. [1]
- Flanders, W. D., Strickland, M. J., and Klein, M. (2017), “A New Method for Partial Correction of Residual Confounding in Time-Series and Other Observational Studies,” *American Journal of Epidemiology*, 185, 941–949. [2]
- Friguet, C., Kloareg, M., and Causeur, D. (2009) “A Factor Model Approach to Multiple Testing Under Dependence,” *Journal of the American Statistical Association*, 104, 1406–1415. [2]
- Gagnon-Bartsch, J. A., and Speed, T. P. (2012), “Using Control Genes to Correct for Unwanted Variation in Microarray Data,” *Biostatistics*, 13, 539–552. [2,12]
- Goldberger, A. S. (1972), “Structural Equation Methods in the Social Sciences,” *Econometrica*, 40, 979–1001. [1]
- Gray, D. L., O’Brien, K. D., D’Alessio, D. A., Brehm, B. J., and Deeg, M. A. (2008), “Plasma Glycosylphosphatidylinositol-Specific Phospholipase D Predicts the Change in Insulin Sensitivity in Response to a Low-Fat but not a Low-Carbohydrate Diet in Obese Women,” *Metabolism*, 57, 473–478. [12]
- Grimmer, J., Knox, D., and Stewart, B. M. (2020), “Naive Regression Requires Weaker Assumptions than Factor Models to Adjust for Multiple Cause Confounding,” arXiv preprint arXiv:2007.12702. [2,3,5,10]
- Guo, Z., Čevid, D., and Bühlmann, P. (2020), “Doubly Debiased Lasso: High-Dimensional Inference Under Hidden Confounding and Measurement Errors,” arXiv preprint arXiv:2004.03758. [2]
- Hu, Y., and Schennach, S. M. (2008), “Instrumental Variable Treatment of Nonclassical Measurement Error Models,” *Econometrica*, 76, 195–216. [4]
- Imai, K., and Jiang, Z. (2019), “Comment: The Challenges of Multiple Causes,” *Journal of the American Statistical Association*, 114, 1605–1610. [2,8]
- Imbens, G. W., and Newey, W. K. (2009), “Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity,” *Econometrica*, 77, 1481–1512. [6]
- Kang, H., Zhang, A., Cai, T. T., and Small, D. S. (2016), “Instrumental Variables Estimation with Some Invalid Instruments and its Application to Mendelian Randomization,” *Journal of the American Statistical Association*, 111, 132–144. [7]
- Kim, J.-O., and Mueller, C. W. (1978), *Factor Analysis: Statistical Methods and Practical Issues*, Beverly Hills, CA: Sage. [8,12]
- Kong, D., Yang, S., and Wang, L. (2022), “Identifiability of Causal Effects with Multiple Causes and a Binary Outcome,” *Biometrika*, 109, 265–272. [2]
- Kress, R. (1989), *Linear Integral Equations*, Berlin: Springer. [5,8]
- Kuroki, M., and Pearl, J. (2014), “Measurement Bias and Effect Restoration in Causal Inference,” *Biometrika*, 101, 423–437. [2,4]
- Leek, J. T., and Storey, J. D. (2007), “Capturing Heterogeneity in Gene Expression Studies by Surrogate Variable Analysis,” *PLoS Genet*, 3, e161. [2]
- Lehman, E., and Scheffe, H. (1950), “Completeness, Similar Regions and Unbiased Tests. Part I,” *Sankhya*, 10, 219–236. [4]
- Lin, W., Feng, R., and Li, H. (2015), “Regularization Methods for High-Dimensional Instrumental Variables Regression with an Application to Genetical Genomics,” *Journal of the American Statistical Association*, 110, 270–288. [12]
- Lipsitch, M., Tchetgen Tchetgen, E., and Cohen, T. (2010), “Negative Controls: A Tool for Detecting Confounding and Bias in Observational Studies,” *Epidemiology*, 21, 383–388. [1]
- Luo, X., and Wei, Y. (2019), “Batch Effects Correction with Unknown Subtypes,” *Journal of the American Statistical Association*, 114, 581–594. [2,4]
- Manski, C. F. (1990), “Nonparametric Bounds on Treatment Effects,” *The American Economic Review*, 80, 319–323. [1]
- McLachlan, G. J., Lee, S. X., and Rathnayake, S. I. (2019), “Finite Mixture Models,” *Annual Review of Statistics and Its Application*, 6, 355–378. [11,13]

- Miao, W., Geng, Z., and Tchetgen Tchetgen, E. (2018), “Identifying Causal Effects with Proxy Variables of an Unmeasured Confounder,” *Biometrika*, 105, 987–993. [1,4,6]
- Miao, W., Shi, X., and Tchetgen Tchetgen, E. (2018), “A Confounding Bridge Approach for Double Negative Control Inference on Causal Effects.” arXiv preprint arXiv:1808.04945. [11]
- Miao, W., and Tchetgen Tchetgen, E. J. (2016), “On Varieties of Doubly Robust Estimators Under Missingness not at Random with a Shadow Variable,” *Biometrika*, 103, 475–482. [4]
- Newey, W. K., and Powell, J. L. (2003), “Instrumental Variable Estimation of Nonparametric Models,” *Econometrica*, 71, 1565–1578. [4,5,6,8]
- Ogburn, E. L., Rotnitzky, A., and Robins, J. M. (2015), “Doubly Robust Estimation of the Local Average Treatment Effect Curve,” *Journal of the Royal Statistical Society, Series B*, 77, 373–396. [6]
- Ogburn, E. L., Shpitser, I., and Tchetgen Tchetgen, E. J. (2019), “Comment on ‘Blessings of Multiple Causes’,” *Journal of the American Statistical Association*, 114, 1611–1615. [2,4]
- (2020), “Counterexamples to ‘The Blessings of Multiple Causes’ by Wang and Blei,” arXiv preprint arXiv:2001.06555. [2,5]
- Ogburn, E. L., and VanderWeele, T. J. (2012), “On the Nondifferential Misclassification of a Binary Confounder,” *Epidemiology*, 23, 433–439. [2]
- Owen, A. B., and Wang, J. (2016), “Bi-Cross-Validation for Factor Analysis,” *Statistical Science*, 31, 119–139. [11,13]
- Pearl, J. (1995), “Causal Diagrams for Empirical Research,” *Biometrika*, 82, 669–688. [3]
- Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D. (2006), “Principal Components Analysis Corrects for Stratification in Genome-Wide Association Studies,” *Nature Genetics*, 38, 904–909. [2]
- Richardson, T. S., and Robins, J. M. (2014), “ACE Bounds; SEMs with Equilibrium Conditions,” *Statistical Science*, 29, 363–366. [1]
- Robins, J. (1986), “A New Approach to Causal Inference in Mortality Studies with a Sustained Exposure Period—Application to Control of the Healthy Worker Survivor Effect,” *Mathematical Modelling*, 7, 1393–1512. [3]
- Robins, J. M. (1994), “Correcting for Non-compliance in Randomized Trials Using Structural Nested Mean Models,” *Communications in Statistics-Theory and Methods*, 23, 2379–2412. [1]
- Rosenbaum, P. R., and Rubin, D. B. (1983), “Assessing Sensitivity to an Unobserved Binary Covariate in an Observational Study with Binary Outcome,” *Journal of the Royal Statistical Society, Series B*, 45, 212–218. [1]
- Rousseeuw, P. J., and Leroy, A. M. (2005), *Robust Regression and Outlier Detection* (Vol. 589), New York: Wiley. [9]
- Schneeberger, M. (2019), “Irx3, a New Leader on Obesity Genetics,” *EBioMedicine*, 39, 19–20. [12]
- Shi, X., Miao, W., Nelson, J. C., and Tchetgen Tchetgen, E. (2020), “Multiply Robust Causal Inference with Double Negative Control Adjustment for Categorical Unmeasured Confounding,” *Journal of the Royal Statistical Society, Series B*, 82, 521–540. [1,6]
- Small, D. S., Tan, Z., Ramsahai, R. R., Lorch, S. A., Brookhart, M. A. (2017), “Instrumental Variable Estimation with a Stochastic Monotonicity Assumption,” *Statistical Science*, 32, 561–579. [1]
- Tchetgen Tchetgen, E. J., Ying, A., Cui, Y., Shi, X., and Miao, W. (2020), “An Introduction to Proximal Causal Learning,” arXiv preprint arXiv:2009.10982. [1,2,4,6]
- Titterton, D. M., Smith, A. F., and Makov, U. E. (1985), *Statistical Analysis of Finite Mixture Distributions*, New York: Wiley. [4,8]
- Wang, J., Zhao, Q., Hastie, T., and Owen, A. B. (2017), “Confounder Adjustment in Multiple Hypothesis Testing,” *The Annals of Statistics*, 45, 1863–1894. [2,4,7,12,13]
- Wang, L., and Tchetgen Tchetgen, E. (2018), “Bounded, Efficient and Multiply Robust Estimation of Average Treatment Effects Using Instrumental Variables,” *Journal of the Royal Statistical Society, Series B*, 80, 531–550. [2,6]
- Wang, S., Yehya, N., Schadt, E. E., Wang, H., Drake, T. A., and Lusis, A. J. (2006), “Genetic and Genomic Analysis of a Fat Mass Trait with Complex Inheritance Reveals Marked Sex Specificity,” *PLoS Genetics*, 2, e15. [11]
- Wang, Y., and Blei, D. M. (2019), “The Blessings of Multiple Causes,” *Journal of the American Statistical Association*, 114, 1574–1596. [2,4,5,10]
- (2020), “Towards Clarifying the Theory of the Deconfounder,” arXiv preprint arXiv:2003.04948. [2]
- Wheatcroft, S. B., Kearney, M. T., Shah, A. M., Ezzat, V. A., Miell, J. R., MODO, M., Williams, S. C., Cawthorn, W. P., Medina-Gomez, G., Vidal-Puig, A., Sethi, J. K., and Crossey, P. A. (2007), “IGF-Binding Protein-2 Protects Against the Development of Obesity and Insulin Resistance,” *Diabetes*, 56, 285–294. [12]
- Wright, P. G. (1928), *Tariff on Animal and Vegetable Oils*, New York: Macmillan. [1]
- Yakowitz, S. J., and Spragins, J. D. (1968), “On the Identifiability of Finite Mixtures,” *The Annals of Mathematical Statistics*, 39, 209–214. [4]